University of Konstanz
Department of Economics

Disagreement, Uncertainty and the True Predictive Density

Fabian Krüger and Ingmar Nolte

Working Paper Series
2011-43

Disagreement, Uncertainty and the True Predictive Density*

Fabian Krüger†
University of Konstanz
CoFE

Ingmar Nolte‡
University of Warwick
FERC, CoFE

September 1, 2011

Abstract

This paper generalizes the discussion about disagreement versus uncertainty in macroeconomic survey data by emphasizing the importance of the (unknown) true predictive density. Using a forecast combination approach, we ask whether cross-sections of survey point forecasts help to approximate the true predictive density. We find that although these cross-sections perform poorly individually, their inclusion into combined predictive densities can significantly improve upon densities relying solely on time series information.

JEL classification: C53, C83, E.7, F.7

Keywords: Disagreement, Uncertainty, Predictive Density, Forecast Combination

---

*We would like to thank Michael Clements, Holger Dette, Nikolaus Hautsch, Sandra Nolte, Winfried Pohlmeier, Ruben Seiberlich and Kenneth Wallis as well as the participants of the faculty seminar at Humboldt Universität zu Berlin, the second Humboldt-Copenhagen Conference in Financial Econometrics, and the 65th European Meeting of the Econometric Society for helpful comments. Skilled research assistance by Horatio Cuesdeanu is gratefully acknowledged. All remaining errors are ours.

†Department of Economics, Box 124, University of Konstanz, 78457 Konstanz, Germany. Phone +49 7531 883753, Fax -4450, email: Fabian.Krueger@uni-konstanz.de. Financial support from the Fritz Thyssen foundation through the project "Analysis, Modeling and Prediction of Multivariate Volatility Processes" is gratefully acknowledged.

‡Warwick Business School, Financial Econometrics Research Centre (FERC), University of Warwick, CV4 7AL, Coventry, United Kingdom. Phone +44-24-765-72838, Fax -23779, email: Ingmar.Nolte@wbs.ac.uk.

1 Introduction

Decision making requires – beyond plain point forecasts – information about the uncertainty surrounding future events.[1] In economics, surveys among experts have been an important source used for constructing measures of uncertainty.

Following the seminal article by Zarnowitz and Lambros (1987), the last decades have witnessed an extensive debate on how to best measure predictive uncertainty from expert surveys. In response to this challenge, the recent literature tends to construct variances from predictive histograms[2] which contain subjective probabilities of the target quantity falling into each of several histogram bins. An additional concept discussed in the literature is "disagreement", computed as the cross-sectional variance of experts' point forecasts. While Mankiw, Reis, and Wolfers (2003) discuss economic implications of disagreement *per se*, a number of studies (e.g. Bomberger (1996), Giordani and Söderlind (2003), Boero, Smith, and Wallis (2008) and Lahiri and Sheng (2010)) analyze whether disagreement can serve as a proxy for uncertainty. Thereby, the benchmark measure of uncertainty is typically constructed from predictive histograms.

Two major assumptions (often implicitly made) underlie the current debate: i) uncertainty is to be measured by second moments and ii) these second moments are best constructed from predictive histograms which are perceived to represent the true predictive density.

Both assumptions are unrealistic. First, abstracting from a pure mean-variance utility concept, alternative uncertainty measures such as quantiles, ranges, number of modi and stochastic dominance considerations are important for forecast users. Second, it is by no means clear (see Giordani and Söderlind (2003)) how to convert predictive histograms into a single variance-based measure of predictive uncertainty. Exemplary issues include the question of whether to take the average of variances constructed from individual-level histograms or the variance of an aggregate histogram, the design of the histogram bins as well as the predictive distribution within each bin.

Moreover, it seems unnecessarily restrictive to rely only on surveys as a single data source, since i) information from historical time series data can readily be added in a forecast combination setting (Wallis (2005)) and ii) the existence of a single superior approximation to

[1] Throughout this paper, and following the literature on macroeconomic survey data, we use the terms "uncertainty" and "risk" synonymously.

[2] For example, predictive histograms are available at the level of individual survey participants in the US-based Survey of Professional Forecasters (SPF). Also, the Bank of England's Survey of External Forecasters publishes an aggregate predictive histogram.

the true data-generating process appears unlikely, especially in the presence of structural breaks (Aiolfi, Capistrán, and Timmermann (2011)). Recent studies by Kascha and Ravazzolo (2010), Jore, Mitchell, and Vahey (2010) and Geweke and Amisano (2011) highlight the success of combining probabilistic forecasts, thereby generalizing findings from the literature on the combination of point forecasts (see Timmermann (2006) for a survey).

This paper analyzes whether the cross-sectional distribution of experts' point forecasts helps to approximate the true predictive densities of several US macroeconomic variables. If this is the case, then cross-sections of survey point forecasts are informative about "predictive uncertainty", in a precise sense and independently of the specific uncertainty measure employed by the forecast user. This question generalizes the debate on "uncertainty" versus "disagreement" along two dimensions: First, rather than focussing on the predictive variance as one specific measure of uncertainty, we consider an entire predictive distribution. Based on this predictive distribution, any desired measure of uncertainty can be constructed. Second, we analyze the information content of the entire cross-sectional distribution of experts' point forecasts, rather than "disagreement" as one specific characteristic of this distribution.

We tackle our research question in a forecast combination setting. Specifically, we construct estimates of the cross-sectional distribution of experts' point forecasts in two distinct survey data sets: The SPF data which contains quantitative forecasts, and the Financial Market Survey administered by the ZEW ("Centre of European Economic Research") containing qualitative forecasts of several US macroeconomic variables. Although qualitative forecasts convey less information than quantitative ones, they may be more reliable as they require less sophistication from survey participants.[3] We ask whether the cross-sectional distributions of point forecasts can add information to predictive densities obtained from three different time series models. These models are specifically chosen to capture a wide range of data sources and functional form assumptions, with the aim of creating a fairly tough benchmark setting for the survey data. We then analyze whether combined predictive densities including survey information lie significantly closer to the true predictive density than combined predictive densities solely relying on time series information.

Closeness to the true predictive density is defined and understood in a Maximum Likelihood sense. Hence, forecasts are evaluated by the log score criterion, which has the property that it is uniquely maximized by the true predictive density. Thus, the goal of finding the true predictive density is equivalent to the maximization of the expected log score.

[3]See Manski (2004) for a careful discussion of topics related to the measurement and interpretation of survey expectations.

The rest of this paper is organized as follows. Section 2 introduces our combination setting, Section 3 presents all individual survey- and time series models, Section 4 presents empirical results, and Section 5 concludes.

2 Model Setup and Data

Let Y_t, $t = 1, \ldots, T$ denote the stationary transform of a macroeconomic variable sampled at quarterly frequency and $\mathcal{F}_t$ the true information process. We are interested in the true two-quarter ahead[4] predictive density $f_t(Y_{t+2}) \equiv f_t(Y_{t+2}|\mathcal{F}_t)$ which is usually unavailable since both the information set $\mathcal{F}_t$ and the true functional form $f_t(\cdot)$ are unknown. What we observe in reality are several incomplete information sets $\mathcal{F}_t^j \subset \mathcal{F}_t, j = 1, \ldots, J$ on which we rely to specify J individual predictive densities $f_t^j(Y_{t+2}) \equiv f_t^j(Y_{t+2}|\mathcal{F}_t^j)$. These may differ in both their underlying information sets and their functional form assumptions; in particular, we will later distinguish between survey- and time series information. In addition, we consider combined predictive densities of the form $f_t^C(Y_{t+2}) \equiv f_t^C(Y_{t+2}|\mathcal{F}_t^1, \ldots, \mathcal{F}_t^J)$.

The specification of a loss function $g(\cdot)$ which expresses the forecast user's utility from the combination of a generic density forecast $\tilde{f}_t(Y_{t+2})$ and an ex-post realized outcome y_{t+2} is essential to our study. A wide range of loss functions have been suggested in the literature; see e.g. Winkler (1996), Gneiting and Raftery (2007) as well as Boero, Smith, and Wallis (2010). We use the log score criterion (Good (1952)) given by $g(y_{t+2}; \tilde{f}_t(Y_{t+2})) = \ln(\tilde{f}_t(y_{t+2}))$ which is the logarithmic value of the predictive density at the ex-post realized outcome. The log score is conceptually related to Maximum Likelihood and the familiar Kullback and Leibler (1951) distance.

The expected log score of a candidate predictive density $\tilde{f}_t(Y_{t+2})$ is given by

$$\mathrm{E}\left[\ln(\tilde{f}_t(Y_{t+2}))\Big|\,\mathcal{F}_t\right] = \int_{\mathbb{R}} \ln(\tilde{f}_t(Y_{t+2}))f_t(Y_{t+2})dY_{t+2}. \tag{1}$$

The Kullback and Leibler (1951) distance between the true predictive density $f_t(Y_{t+2})$ and its approximation $\tilde{f}_t(Y_{t+2})$ is given by

[4]We consider two-quarter ahead predictions since this forecast horizon is covered by both the SPF- and ZEW surveys.

$$\mathrm{KL}(f_t, \tilde{f}_t) = \mathrm{E}\left[\ln\left[\frac{f_t(Y_{t+2})}{\tilde{f}_t(Y_{t+2})}\right]\middle| \mathcal{F}_t\right] \tag{2}$$

$$= \int_{\mathbb{R}} \ln(f_t(Y_{t+2}))f_t(Y_{t+2})dY_{t+2} - \int_{\mathbb{R}} \ln(\tilde{f}_t(Y_{t+2}))f_t(Y_{t+2})dY_{t+2}. \tag{3}$$

Since the first term in (3) does not depend on $\tilde{f}_t$, it is irrelevant for the task of choosing a good predictive density. Hence, maximizing the expected log score is tantamount to minimizing the Kullback and Leibler (1951) distance to the unknown true predictive density $f_t(Y_{t+2})$. The (unique) minimum of $\mathrm{KL}(f_t, \tilde{f}_t)$ (and hence, the unique maximum of the expected log score) is attained by setting $\tilde{f}_t(\cdot) = f_t(\cdot)$; in this case, $\mathrm{KL}(f_t, f_t) = 0$.[5] This establishes that the log score is a "proper" scoring rule: a forecaster wishing to maximize the expected log score cannot do better than revealing what he thinks is the true predictive density $f_t(\cdot)$ (Winkler (1969)).

Clearly, the expected log score in (1) is unobservable in practice. Instead, the predictive density f_t^j is commonly evaluated on the basis of the realized log scores $\{\ln(f_t^j(y_{t+2}))\}_{t=T_c}^{T-2}$ corresponding to the evaluation sample $y_{T_c+2}, ..., y_T$ defined below, where $T_c < T$. The negative of the log score, $-\ln(f_t^j(y_{t+2}))$, is the loss of model f^j at time $t+2$. The corresponding sequence of loss differentials between model f^j and a competing model f^k is given by $\{d_{t+2}^{j,k}\}_{t=T_c}^{T-2}$, where

$$d_{t+2}^{j,k} \equiv \ln(f_t^k(y_{t+2})) - \ln(f_t^j(y_{t+2})). \tag{4}$$

Such sequences of loss differentials directly allow for statistical comparisons of the predictive accuracy of two or more competing models via tests in the spirit of Diebold and Mariano (1995, henceforth, DM) and Hansen (2005), respectively; see Kascha and Ravazzolo (2010) and Bao, Lee, and Saltoglu (2007) for two exemplary applications.

Our empirical study uses information sets of different origins; we specify five models to estimate predictive densities on the basis of these information sets. The first two models are based on survey information, while the last three models are based on time series information. In our analysis, the latter models will serve (individually and in combined form) as benchmark predictive densities. The question we address is whether they can be significantly improved upon via combination with survey information. Our choice of time series based predictive densities is guided by the idea of spanning a wide range of data

[5]See Rubinstein and Kroese (2008, p.31).

sources and functional form assumptions. This should render it fairly tough for the survey based densities to add further information.

The first model is developed around the SPF currently administered by the Federal Reserve Bank of Philadelphia. The survey provides two-quarter ahead point forecasts of important macroeconomic aggregates, at the individual forecaster level (roughly 30-40 participants per period).[6] We employ a nonparametric estimate of the cross-sectional distribution of point forecasts to obtain a predictive density. The second model exploits the ZEW Financial Market Survey which contains individual-level forecasts of roughly 300 finance professionals in qualitative form.[7] We use the Carlson and Parkin (1975) quantification method to estimate the cross-sectional distribution of point forecasts which we again interpret as a predictive density. The third model is based on past observations of the target variable Y_{t+2}. We construct a predictive density from a nonparametric estimate of the distribution of Y_{t+2}, conditional on the single predictor Y_t. Models number four and five are based on a rich set of approximately 100 macroeconomic predictors, in addition to observations from the target variable. Model four is based on the idea of constructing a density forecast for Y_{t+2} from forecasts of a number of different conditional quantiles of Y_{t+2} (Koenker (2005)). As predictors in each of the conditional quantile regressions, we use the most recent value Y_t of the target variable as well as principal components summarizing the macroeconomic predictors at time t (Stock and Watson (2002)). Model five uses the same set of regressors to construct a point forecast of Y_{t+2} and uses the assumption that forecast errors are normally distributed. We provide a more detailed description of all models in the next section.

We consider predictive densities for four quarterly macroeconomic aggregates from the US: The annualized growth rate of real GDP, the annualized CPI inflation rate, the three-month TBILL rate, and the ten-year TBOND rate.[8] The data we use range from 1964/4 to 2009/4; the sample paths of the four variables during this time span are depicted in Figure 1 below. In order to mimic the process of producing and combining forecasts in real time, we split our data into three subsamples: First, observations until T_e ("estimation sample") are used to estimate the parameters of the individual predictive densities. Second, observations between $T_e + 2$ and T_c ("combination sample") are used to estimate unknown parameters

[6]See http://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/ for detailed information about the SPF.

[7]Nolte and Pohlmeier (2007) and Nolte, Nolte, and Pohlmeier (2010) provide detailed data descriptions.

[8]All data were downloaded from the FRED database administered by the Federal Reserve of St. Louis. We compute the GDP growth rate as $Y_t = \ln(X_t) - \ln(X_{t-1})$, where X_t is the quarterly level of real GDP. Finally, we annualize this growth rate. For inflation, we first compute quarterly levels of the CPI index by averaging across three monthly observations. We then compute annualized quarterly growth rates as described for GDP above. For both interest rates, we obtain quarterly levels by averaging over the rates corresponding to all working days during the quarter.

of the combined predictive densities. Third, all combined and individual out-of sample density forecasts are finally evaluated using observations between T_c+2 and T ("evaluation sample"). We initially set T_e to 1992/1 and T_c to 1999/4.[9] We then shift both T_e and T_c in a rolling window fashion, such that the estimation sample always contains $R = 110$ observations and the combination sample always contains $W = 30$ observations. At the end of our forecasting exercise, we have thus produced 39 combined out-of sample density forecasts for observations occurring between 2000/2 and 2009/4. These forecasts form the basis for our comparison of the (individual and combined) models' relative predictive performance.

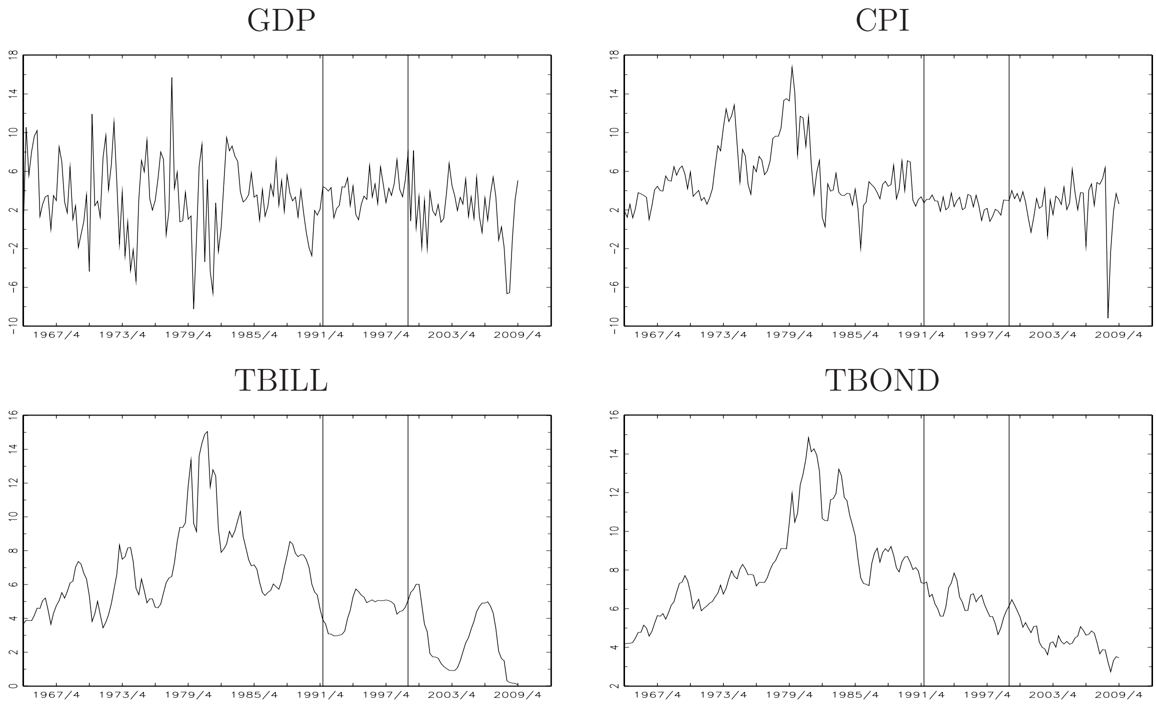

Figure 1: Sample paths of the annualized growth rate of real GDP, the annualized CPI inflation rate, the three-month TBILL rate and the ten-year TBOND rate between 1964/4 and 2009/4. The left vertical line marks 1992/1, the end of our estimation sample in the first forecast recursion. The right vertical line marks 1999/4, the end of our combination sample in the first forecast recursion.

[9]Our choice of T_e is determined by the availability of the ZEW forecasts.

3 Individual Predictive Densities

3.1 Approaches Based on Survey Data

The first two approaches are based on the idea of interpreting the (estimated) *cross-sectional* distribution of point forecasts of Y_{t+2} among a specific group of experts as an approximation to the true predictive density $f_t(Y_{t+2}|\mathcal{F}_t)$. We implement two distinct variants which rely on two different surveys: First, a nonparametric estimate of the cross-sectional distribution of forecasts among SPF participants. Second, a parametric quantification method based estimate for the qualitative forecasts of the ZEW Financial Market Survey.

Model 1: Survey forecast based on SPF data

Let y^1_{it+2} be the point prediction expressed by the ith SPF participant in period t, with $i \in \{1, ..., N^1_t\}$. Superindex "1" expresses that a quantity refers to model 1; similar notation is used in the following whenever an analogous quantity appears in several models. We neglect the identities of the forecasters and view the N^1_t different forecasts as independent draws from the cross-sectional distribution of forecasts of Y_{t+2} based on time t; this distribution can be thought of as representing the range of expectations about Y_{t+2} among the SPF participants. We then use a nonparametric estimate of this distribution as a predictive density.

Formally, we have

$$f^1_t(Y_{t+2}) = \frac{1}{N^1_t h^1_t} \sum_{i=1}^{N^1_t} K\left(\frac{Y_{t+2} - y^1_{it+2}}{h^1_t}\right),$$

where h^1_t is the bandwidth and $K(\cdot)$ is the kernel used for our nonparametric estimate of the cross-sectional distribution of expert forecasts. We employ a Gaussian kernel and choose the bandwidth h^1_t by the rule of thumb due to Silverman (1986).

Model 2: Survey forecast based on ZEW data

Unlike the SPF forecasts, the ZEW forecasts are qualitative. Rather than a quantitative prediction y^2_{it+2}, we thus observe three dummy variables $(u_{it+2}, s_{it+2}, d_{it+2})$ ("up/ same/ down") which code the forecast of the ith survey participant, with $i \in \{1, ..., N^2_t\}$. The Carlson and Parkin (1975) method[10] assumes the following relationship between latent

[10]See Pesaran and Weale (2006) for a discussion and Nolte and Pohlmeier (2007) for an application to the ZEW data.

continuous and observed directional quantities:

$$(u_{it+2}, s_{it+2}, d_{it+2}) \equiv \begin{cases} (1,0,0) & y^2_{it+2} \geq \lambda_{ut+2} \\ (0,1,0) & \lambda_{dt+2} \leq y^2_{it+2} < \lambda_{ut+2} \\ (0,0,1) & y^2_{it+2} < \lambda_{dt+2} \end{cases},$$

where $\lambda_{dt+2} < \lambda_{ut+2}$ are the respective down and up threshold series.

Carlson and Parkin (1975) assume that the cross-section of latent quantitative forecasts made at time t is drawn from a normal distribution: $y^2_{it+2} \sim N(\mu_{t+2}, \sigma^2_{t+2})$. Computing the individual level "up" and "down" probabilities and replacing them by their sample counterparts yields:

$$u_{t+2} = 1 - \Phi\left(\frac{\lambda_{ut+2} - \mu_{t+2}}{\sigma_{t+2}}\right), \tag{5}$$

$$d_{t+2} = \Phi\left(\frac{\lambda_{dt+2} - \mu_{t+2}}{\sigma_{t+2}}\right), \tag{6}$$

where $u_{t+2} \equiv \frac{1}{N^2_t}\sum_{i=1}^{N^2_t} u_{it+2}$ and $d_{t+2} \equiv \frac{1}{N^2_t}\sum_{i=1}^{N^2_t} d_{it+2}$ denote the cross-sectional shares of "up" and "down" forecasts recorded at time t, N^2_t is the corresponding number of micro-level forecasts and $\Phi(\cdot)$ denotes the cumulative density function (cdf) of the standard normal distribution.

$(\mu_{t+2}, \sigma^2_{t+2})$ are identified from (5) and (6) only under the assumption that the thresholds λ_{dt+2} and λ_{ut+2} are known. Therefore, we use threshold series based on individual-level responses to an additional questionnaire sent out by the ZEW from time to time. Having estimated μ_{t+2} and σ^2_{t+2} in this way, we construct an estimate of the cross-sectional distribution of point forecasts among the ZEW survey participants, based on the Carlson and Parkin (1975) assumptions.

3.2 Approaches Based on Time Series Data

In addition to the survey-based predictive densities we consider three different approaches based on time series data.

Model 3: Nonparametric conditional density estimation

Our third predictive density is a nonparametric estimate of the conditional distribution of Y_{t+2} given Y_t, evaluated at the most recently observed value y_t.[11] Formally, we have

$$\begin{aligned} f_t^3(Y_{t+2}) &= \hat{f}\left(Y_{t+2}|Y_t = y_t\right) = \frac{\hat{f}(Y_{t+2}, y_t)}{\hat{f}(y_t)}, \\ \hat{f}(Y_{t+2}, Y_t) &= \frac{1}{R(h_t^3)^2} \sum_{j=0}^{R-1} K\left(\frac{Y_{t+2} - y_{t-j}}{h_t^3}\right) K\left(\frac{Y_t - y_{t-j-2}}{h_t^3}\right), \\ \hat{f}(Y_t) &= \frac{1}{Rh_t^3} \sum_{j=0}^{R-1} K\left(\frac{Y_t - y_{t-j}}{h_t^3}\right). \end{aligned} \tag{7}$$

Estimation of $f_t^3(Y_{t+2})$ is performed using a rolling window of $R = 110$ quarterly observations. As for Model 1, we use a Gaussian kernel $K(\cdot)$; we select the bandwidth h_t^3 via Scott's rule (Härdle, Müller, Sperlich, and Werwatz (2004, p.73)).

Model 4: Quantile regression

Our fourth predictive density is based on two-step ahead forecasts $q_{\alpha t}(Y_{t+2})$ of the α quantile of Y_{t+2}:[12]

$$q_{\alpha t}(Y_{t+2}) = \hat{\beta}_{\alpha 0} + \hat{\beta}_{\alpha 1} y_t + \hat{\beta}_{\alpha 2} pc_t, \tag{8}$$

where $\alpha \in (0, 1)$ and pc_t is the first principal component extracted from a set of 92 stationary macroeconomic predictors; see the Appendix for a description of all underlying variables and their transformations and Stock and Watson (2002) for a classic reference on macroeconomic forecasting using principal components.

The estimated parameter vector $\hat{\beta}_\alpha = \left[\hat{\beta}_{\alpha 0}, \hat{\beta}_{\alpha 1}, \hat{\beta}_{\alpha 2}\right]'$ in (8) is given by

$$\hat{\beta}_\alpha = \underset{b \in \mathbb{R}^3}{\operatorname{argmin}} \sum_{j=0}^{R-1} \left(y_{t-j} - x'_{t-j-2} b\right) \left(\alpha - \mathbf{1}\left(y_{t-j} - x'_{t-j-2} b < 0\right)\right), \tag{9}$$

[11] See Härdle, Müller, Sperlich, and Werwatz (2004, Section 3.6) for a textbook treatment of multivariate density estimation.

[12] The idea of constructing a predictive density from quantile regressions has been pursued by Cenesizoglu and Timmermann (2008) and Coroneo and Veredas (2010). See Komunjer (2005) for a treatment of the statistical properties of regression quantiles in a time series context.

where $x_t = \begin{bmatrix} 1 & y_t & pc_t \end{bmatrix}'$ and $\mathbf{1}(\cdot)$ is the indicator function. As for Model 3 above, we use a rolling window of $R = 110$ quarterly observations for parameter estimation and construction of the principal component pc_t.

In principle, we could estimate quantile regressions for a fine grid of levels α and construct a predictive histogram directly from the resulting predictions $q_{\alpha t}(Y_{t+2})$ in (8). However, this approach would be problematic for a number of reasons: First, the predicted quantiles $q_{\alpha t}(Y_{t+2})$ do not necessarily satisfy the logical requirement of monotonicity in α, especially if we consider a fine grid of values for α.[13] Second, since the different quantile levels are treated in isolation, the predictions $q_{\alpha t}(Y_{t+2})$ are an implausibly rough function of α. Third, the predicted "tail quantiles" (α near zero or one) are very unreliable due to our small sample size typical of macroeconomic time series.

In order to resolve the first two problems, we proceed as follows: We first obtain predicted quantiles $q_{\alpha t}(Y_{t+2})$ for a fine grid of values $\alpha \in \{0.005, 0.01, ..., 0.995\}$. We then run a local linear regression of $q_{\alpha t}(Y_{t+2})$ on the quantile level α, subject to the constraint that the resulting prediction $\tilde{q}_{\alpha t}(Y_{t+2})$ be strictly increasing in α. We implement the procedure of Dette, Neumeyer, and Pilz (2006) for this purpose. This provides us with a new sequence of predicted quantiles $\tilde{q}_{\alpha t}(Y_{t+2})$ which is both smooth and monotone in α. In order to resolve the third problem, we impose normality on the predicted quantiles at levels α smaller than 0.05. This is achieved by equalizing these quantiles to the quantiles of a normally distributed variable with mean $\tilde{q}_{0.5t}(Y_{t+2})$ and standard deviation chosen to match $\tilde{q}_{0.05t}(Y_{t+2})$. We proceed analogously for quantiles at levels α exceeding 0.95.

To summarize, our transformed quantile predictions $q^*_{\alpha t}(Y_{t+2})$ are given by

$$q^*_{\alpha t}(Y_{t+2}) = \begin{cases} \tilde{q}_{0.5t}(Y_{t+2}) + \frac{\tilde{q}_{0.05t}(Y_{t+2}) - \tilde{q}_{0.5t}(Y_{t+2})}{\Phi^{-1}(0.05)} \Phi^{-1}(\alpha) & \alpha < 0.05 \\ \tilde{q}_{\alpha t}(Y_{t+2}) & \alpha \in [0.05, 0.95] \\ \tilde{q}_{0.5t}(Y_{t+2}) + \frac{\tilde{q}_{0.95t}(Y_{t+2}) - \tilde{q}_{0.5t}(Y_{t+2})}{\Phi^{-1}(0.95)} \Phi^{-1}(\alpha) & \alpha > 0.95 \end{cases},$$

where $\Phi^{-1}(\cdot)$ denotes the inverse of the cdf of the standard normal distribution. We finally obtain the predictive density $f^4_t(Y_{t+2})$ by constructing a histogram from the transformed quantile predictions $q^*_{\alpha t}(Y_{t+2})$ at levels $\alpha \in \{0.05, 0.1, \ldots, 0.95\}$ and imposing normality on the tails as discussed above.

[13]This phenomenon, which is often referred to as "quantile crossing", is well known in the literature; see e.g. Dette and Volgushev (2008) and the references therein.

Model 5: Parametric distribution around a mean forecast

For the fifth predictive density, we construct a parametric mean forecast and then impose a specific distributional assumption (normality) on the prediction errors. Specifically, we have

$$\begin{aligned}\hat{\mu}_{t+2} &= \hat{\gamma}_0 + \hat{\gamma}_1 y_t + \hat{\gamma}_2 pc_t, \\ \hat{\sigma}_{t+2} &= \sqrt{\frac{1}{R-1}\sum_{j=0}^{R-1}\left(y_{t-j} - x'_{t-j-2}\hat{\gamma}\right)^2}, \\ f_t^5(Y_{t+2}) &= \frac{1}{\hat{\sigma}_{t+2}}\phi\left(\frac{Y_{t+2}-\hat{\mu}_{t+2}}{\hat{\sigma}_{t+2}}\right).\end{aligned}$$

where $x_t = \begin{bmatrix}1 & y_t & pc_t\end{bmatrix}'$, $\hat{\gamma} = \begin{bmatrix}\hat{\gamma}_0 & \hat{\gamma}_1 & \hat{\gamma}_2\end{bmatrix}'$, $\phi(\cdot)$ denotes the probability distribution function of the standard normal distribution and pc_t is as in Model 4 above. The parameter vector $\hat{\gamma}$ is estimated via OLS using a rolling window of $R = 110$ observations.

3.3 Discussion and Forecast Combinations

The five models we consider differ with respect to both their underlying information sets and their functional form assumptions. This causes them to produce very different predictive distributions, in terms of location, dispersion, skewness, kurtosis and shape. While models two and five rest on restrictive normality assumptions, the other three models can generate asymmetric, fat-tailed and/or multimodal densities. Figures 2 and 3 display examples of all five predictive distributions, for i) the TBILL rate in the fourth quarter of 2000 and ii) the CPI inflation rate during the third quarter of 2008.

In addition to analyzing individual predictive densities, we consider forecast combinations as a natural next step to approximate the true predictive density. Moreover, to address the question whether the survey based densities contain incremental information, we will later consider combinations among different sets of models (time series information only versus time series- and survey information). Combination of point forecasts has a long and successful tradition in economics; see Timmermann (2006) for a survey. Combination of predictive densities has recently been pursued by Hall and Mitchell (2007), Geweke and Amisano (2011), Kascha and Ravazzolo (2010) and Jore, Mitchell, and Vahey (2010). In our study the individual predictive densities to be combined are very heterogeneous, spanning a wide range of data sources and functional form assumptions. We consider a number of combination approaches; all of them are based on the idea of specifying a mixture density

$$\begin{aligned} f_t^C(Y_{t+2}) &= \sum_{j=1}^{J} w_t^j f_t^j(Y_{t+2}), \\ w_t^j &\in [0,1], \\ \sum_{j=1}^{J} w_t^j &= 1, \end{aligned}$$

to combine the individual predictive densities (Wallis (2005)). We consider four standard ways of specifying the weights w_t^j (see Kascha and Ravazzolo (2010)):

- *Equal weights (E):*
$$w_t^j = \frac{1}{J}.$$

- *Recursive log score weights (RLS):*
$$w_t^j = \frac{\exp(\sum_{l=0}^{W-1} \ln(f_{t-2-l}^j(y_{t-l})))}{\sum_{j=1}^{J} \exp(\sum_{l=0}^{W-1} \ln(f_{t-2-l}^j(y_{t-l})))},$$
i.e. weights are chosen in proportion to the different models' track record during the last W periods. We set $W = 30$ in the following.

- *Recursive best model (RB):*
$$w_t^j = \begin{cases} 1 & \text{if } j = \underset{k}{\operatorname{argmax}} \sum_{l=0}^{W-1} \ln(f_{t-2-l}^k(y_{t-l})) \\ 0 & \text{else} \end{cases},$$
i.e. the model with the best track record during the last W observations is selected.

- *Optimal in-sample weights (OIS):*
$$w_t^j = \underset{\omega^1,\ldots,\omega^J}{\operatorname{argmax}} \sum_{l=0}^{W-1} \ln(\sum_{j=1}^{J} \omega^j f_{t-2-l}^j(y_{t-l})),$$
under the constraint that the ω^j are positive and sum to unity. This scheme amounts to a numerical search for the weight vector which maximizes the average log score for the last W observations.[14]

[14]See Hall and Mitchell (2007) who originally proposed this combination scheme, and Geweke and Amisano (2011) who provide a formal discussion.

Note that the equal weights combination scheme provides insurance against idiosyncratic model failure. This is particularly effective if the predictive densities are heterogeneous, so that simultaneous failure of all models is unlikely. By contrast, the second and third combination schemes aim at dynamically switching between individual models, such as to emphasize successful over less successful predictive densities. Thereby, recursive log score weights constitute a less aggressive switching mechanism than the recursive best model selector. These schemes are promising if relative model performance is persistent, so that past relative performance is a good proxy for future relative performance. Optimal in-sample weights can, in principle, produce both balanced ($w_t^j \approx \frac{1}{J}$) and unbalanced combination weights, depending on what performed better in the past. Once again, however, some degree of stability in relative model performance is required to justify the implicit notion that historically successful combination weights will perform well in the future.

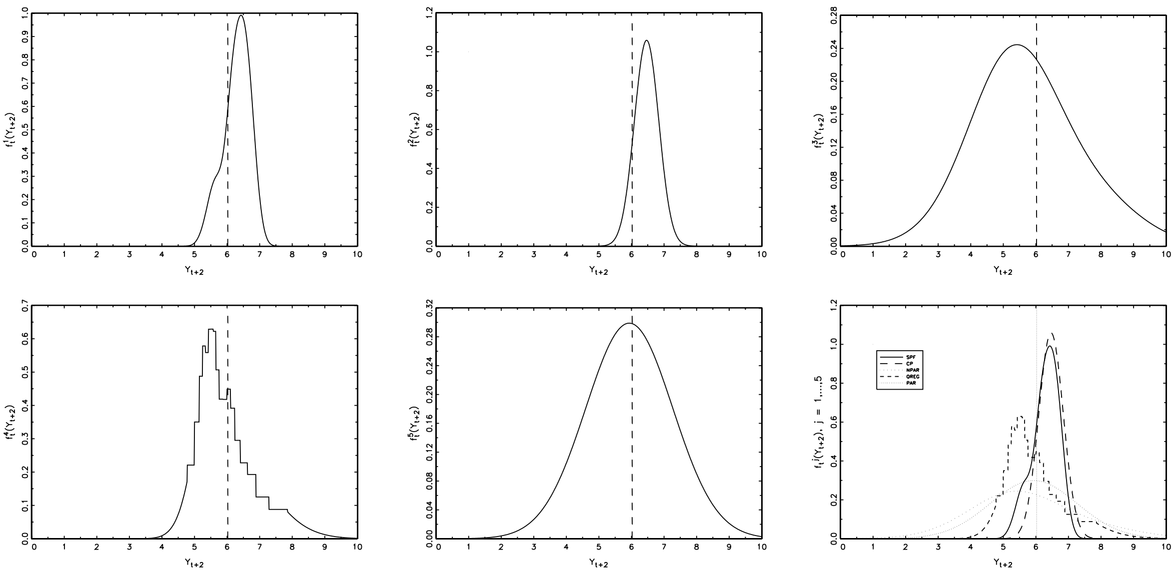

Figure 2: Comparison of five predictive densities $f_t^j(Y_{t+2}), j = 1, \ldots, 5$, *with* Y_{t+2} *representing the TBILL rate in the fourth quarter of 2000.* The first row displays the survey-based predictive densities constructed from the SPF/ZEW data ($j = 1, 2$) as well as the nonparametric predictive density ($j = 3$). The second row displays the quantile regression based- and parametric predictive densities ($j = 4, 5$), as well as a joint graph of all five alternatives. The vertical line marks y_{t+2}, the TBILL rate which actually materialized in 2000/4.

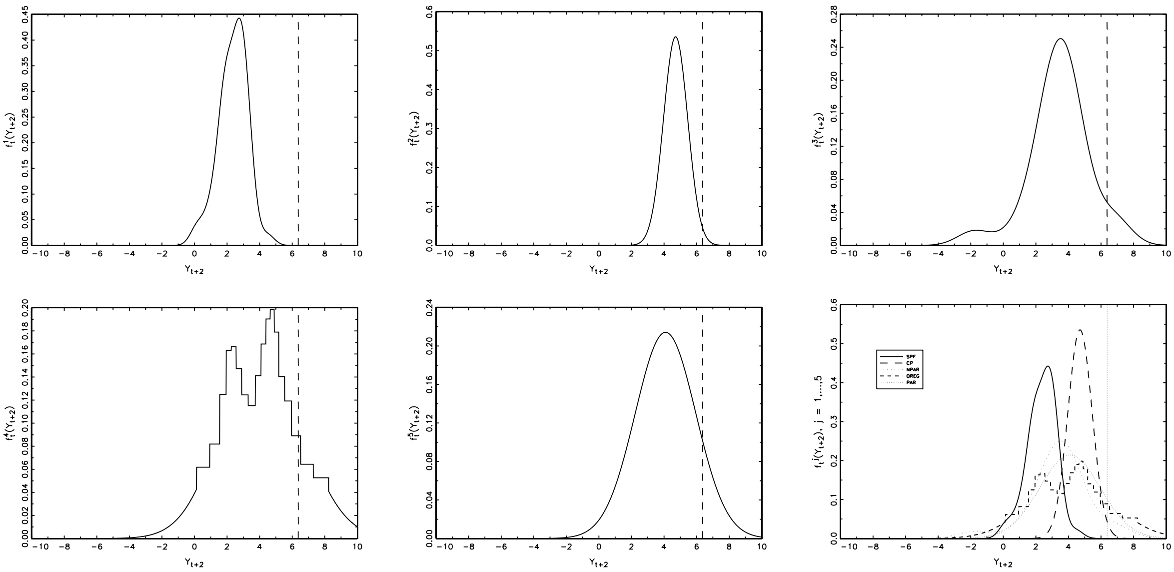

Figure 3: Comparison of five predictive densities $f_t^j(Y_{t+2}), j = 1, \ldots, 5$, *with* Y_{t+2} *representing the annualized CPI inflation rate during the third quarter of 2008.* The first row displays the survey-based predictive densities constructed from the SPF/ZEW data ($j = 1, 2$) as well as the nonparametric predictive density ($j = 3$). The second row displays the quantile regression based- and parametric predictive densities ($j = 4, 5$), as well as a joint graph of all five alternatives. The vertical line marks y_{t+2}, the annualized CPI inflation rate which actually materialized in 2008/3.

4 Empirical Results

The log scores for all five individual predictive densities and the four different weighting schemes are presented in Table 1. The associated scatter plots[15] are depicted in Figure 4.

A first important observation is that the equally weighted mixture combination performs very well relative to all other individual and combined predictive densities. For the CPI inflation- and TBILL rate series, the equally weighted combination scheme outperforms all competitors in terms of average log score over the evaluation period. For the two other series, it performs only marginally worse than the best competitor. In terms of the Superior Predictive Ability (SPA) test by Hansen (2005), there is no evidence that the equally weighted scheme is dominated by a competitor at any conventional level of significance; this is true for all four time series. The performance of the other three combination schemes is somewhat instable across the four series. While OIS weights perform quite satisfactory, both recursive weighting schemes (RLS and RB) yield considerably worse results. This suggests that relative model performance can hardly be predicted. Our finding that simple equal weights perform quite well mirrors a stylized fact from the literature on combinations of point forecasts (the "forecast combination puzzle"), that simple averages across all pre-

[15]We omit the three weighting schemes other than equal weights for ease of presentation.

dictions are often superior to more sophisticated specifications of the weights assigned to the individual forecasts; see Jose and Winkler (2008).[16] In this literature, Smith and Wallis (2009) suggest that involved specifications of the combination weights produce estimation noise which increases the variance of the resulting combined forecast, to an extent which dominates potential bias reductions through flexible weights. Our results, as well as results by Geweke and Amisano (2011), suggest that the "forecast combination puzzle" seems to apply also to combinations of predictive densities.[17] In the light of these results, we focus on the equally weighted combination scheme in the following.

	GDP growth		CPI inflation		TBILL		TBOND	
	MLS	SPA	MLS	SPA	MLS	SPA	MLS	SPA
SPF	-7.11	1.91	-7.95	7.54	-6.62	6.06	-1.97	12.41
CP	-30.14	0.09	-23.05	0.03	-4.91	1.74	-1.26	20.89
NPAR	-2.59	79.54	-2.94	54.17	-1.84	0.00	-1.60	0.00
QREG	-2.60	62.26	-2.79	28.25	-1.85	8.58	-1.91	14.88
PAR	-2.69	6.61	-2.97	14.52	-1.60	6.90	-1.08	57.95
E	-2.59	70.20	-2.49	99.34	-1.20	83.73	-1.02	96.94
RLS	-2.58	97.17	-2.90	44.69	-1.88	7.40	-1.09	46.45
RB	-2.62	31.37	-3.07	19.67	-1.91	9.87	-1.08	64.72
OIS	-2.61	72.40	-2.78	41.20	-1.29	32.66	-1.02	95.11

Table 1: Results of two-step ahead density forecasts for the evaluation period 2000/2 to 2009/4 (39 quarterly data points). Mean log scores (MLS; first column) are defined as MLS $\equiv \frac{1}{39}\sum_{t=T_c}^{T-2}\ln(\tilde{f}_t(y_{t+2}))$, where $\tilde{f}_t(\cdot)$ is a generic two-step ahead predictive density. T_c and T correspond to 1999/4 and 2009/2, respectively. P-values of the Superior Predictive Ability test of Hansen (2005) (SPA; second column; values in percent) refer to the null hypothesis that a particular model is not dominated by any competitor. The relevant loss function is the negative of the log score. Following Hansen and Lunde (2005), we use a block length parameter of $q = 0.5$ and a sample size of $B = 10000$ in our bootstrap implementation.

[16]Interestingly, DeMiguel, Garlappi, and Uppal (2009) obtain very similar results in the context of choosing portfolio weights.

[17]Geweke and Amisano (2011, p.10) report that even with hundreds of daily stock return observations, it is challenging to beat a simple equally weighted mixture model.

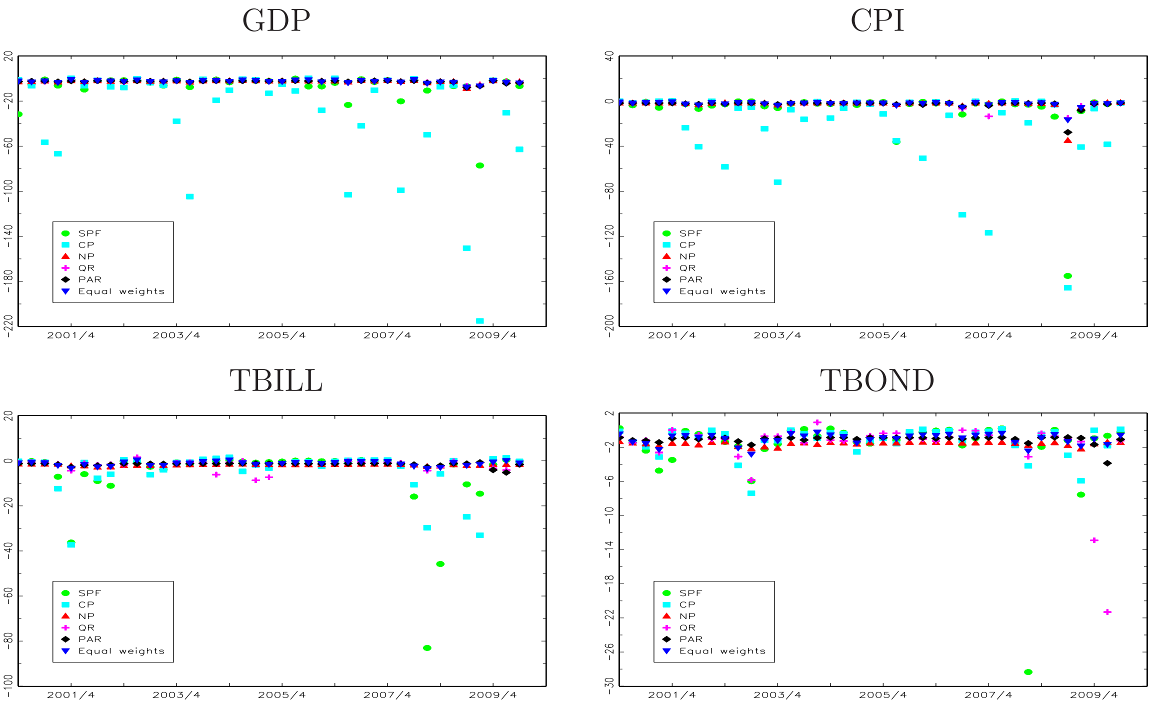

Figure 4: Scatter plots of log scores for two-quarter ahead predictive densities: Annualized growth rate of real GDP, annualized CPI inflation rate, ten-year TBOND rate and three-month TBILL rate. The evaluation period ranges from 2000/2 – 2009/4 (39 quarterly data points).

A second important observation is that the individual survey models perform poorly in terms of the log score. The two survey-based predictive densities are clearly inferior to the time series specifications we consider. For the ZEW-based density, the null hypothesis of the SPA test is rejected at the 5% level for all time series except TBOND. Similarly, for the SPF-based density and all series except TBOND, the SPA null hypothesis is rejected at the 5 % or 10 % levels. The poor *average* performance of the two survey densities is due to the fact that they are too narrow; this causes some realizations to fall far into the tails of their support which results in very low values of the log score criterion (again see Figure 4). For most points in our evaluation sample, one of the two survey densities constitutes the worst model (see Table 2). Thus in general, the estimated cross-sectional distribution of point forecasts appears to be an inappropriate predictive distribution. This confirms and generalizes the findings of Lahiri and Sheng (2010) and Boero, Smith, and Wallis (2008) who show that cross-sectional disagreement tends to understate (their measures of) predictive uncertainty. Note, however, that Bomberger (1996) suggests that disagreement tracks uncertainty only up to a factor of proportionality.

	GDP growth		CPI inflation		TBILL		TBOND	
	% best	% worst	% best	% worst	% best	% worst	% best	% worst
SPF	23.1	17.9	23.1	30.8	28.2	15.4	30.8	25.6
CP	28.2	59.0	28.2	59.0	35.9	23.1	17.9	15.4
NPAR	28.2	0.0	35.6	0.0	5.1	46.2	0.0	53.8
QREG	7.7	5.1	12.8	2.6	15.4	10.3	28.2	5.1
PAR	12.8	17.9	10.3	7.7	15.4	5.1	23.1	0.0

Table 2: Relative performance of the five individual predictive densities during our evaluation period 2000/2 to 2009/4 (39 quarterly data points): "% best" denotes the share among 39 evaluation points for which a particular density achieved the highest log score, and analogously for "% worst".

A third important observation is that the individual survey models, although they perform poorly in terms of the log score and often constitute the worst models, are also the best forecasting models for a considerable share of evaluation points (between 17.9% and 35.9%; see Table 2). This suggests that the estimated cross-sectional distributions of point forecasts may still contain valuable information, although they are exceedingly risky when used individually.

	GDP growth		CPI inflation		TBILL		TBOND	
	DMLS	DM stat	DMLS	DM stat	DMLS	DM stat	DMLS	DM stat
TS vs. (TS + SPF + CP)	-0.00	-0.08	-0.01	-0.16	-0.30	-1.98*	-0.15	-2.24*
TS vs. (TS + SPF)	-0.00	-0.04	0.00	-0.17	-0.17	-1.80*	-0.11	-2.22*
TS vs. (TS + CP)	0.01	0.13	0.02	0.27	-0.28	-2.12*	-0.11	-2.59**

Table 3: Comparisons of equally weighted density combinations with- and without survey information during our evaluation period 2000/2 to 2009/4 (39 quarterly data points). "DMLS" denotes the mean log score of combination A minus the mean log score of combination B. "DM stat" gives the Diebold-Mariano test statistic corresponding to the null hypothesis that including survey information improves the expected log score of the combination. The test statistic is computed from an auxiliary regression of the log score differential on a constant, using HAC standard errors. The t-statistic associated with the constant yields the Diebold-Mariano test statistic. One- and two stars indicate significance at the five percent- and one percent levels (one-sided tests).

In Table 3 we report our main results of whether or not the inclusion of the survey based densities significantly improves the log score criterion and hence helps to construct a predictive density that is closer to the true one. We differentiate between including both– or either of the two survey based densities to the pool of three time series based densities. Throughout, we focus on equally weighted combinations of all involved models. We report DM test statistics for mixture combinations with- and without the survey information in Table 3. For the GDP growth- and CPI inflation series, including or excluding the survey based densities does not make a significant difference. In contrast, including (either one or both of) the survey densities significantly improves upon an equally weighted pool of the three time series models for the TBILL– and TBOND series. The corresponding DM test statistics are significant at the 5% and 1% levels (one-sided tests).

These results suggest that although the cross-sectional distributions of point forecasts perform very poorly individually, they contain substantial information which can be exploited, for example via simple equally weighted combinations. Hence, suggestions to dismiss information in cross-sections of point forecasts altogether (e.g. Engelberg, Manski, and Williams (2009)) might be overhasty. This is particularly true since the existing literature tends to focus on disagreement as one specific characteristic of the cross-sectional distribution of point forecasts. The more general question "What is and how do we construct the true predictive density?" has rarely been asked.

5 Conclusion

Measures of disagreement and predictive uncertainty prevalent in the literature are subject to fundamental statistical critique. Both conceptual and practical issues arise. We suggest that the focus of attention should be shifted to the true predictive density, which would naturally overcome the above ambiguities. Hence we present a combination approach in which models relying on distinct information sets and functional form assumptions are combined to approximate the true predictive density. Within this framework we ask the important question whether cross-sections of survey point forecasts reveal information about this density. This question generalizes the debate about disagreement versus uncertainty.

We consider cross-sectional distributions of survey point forecasts from the SPF and the ZEW for GDP growth, inflation, the TBILL rate and the TBOND rate in the US. Individually, both distributions perform poorly for all variables. Nevertheless, we show that their inclusion significantly improves the quality of combined predictive densities for the TBILL and TBOND rates while it does not affect the quality of the combination for GDP growth and CPI inflation. These results suggest that information in cross-sections of point forecasts should not be excluded a priori when considering measures of predictive uncertainty.

More generally, we find that combining predictive densities is a successful strategy in that combinations can considerably improve upon all of their components. The stable combination schemes we consider (in particular, equal weights) effectively exploit information from predictive densities which are inappropriate when considered in isolation. In contrast, we find little support for combination mechanisms which aim at recursively selecting the best individual models. This suggests that relative model performance is hard to predict in our application. However, further research is needed in order to fully understand the nature of optimal combinations of predictive densities. Work along the lines of Geweke and Amisano (2011) and Clements and Harvey (2011), who consider the case of a binary response variable, promises to yield important insights in this respect.

References

Aiolfi, M., C. Capistrán, and A. Timmermann (2011): “Forecast Combinations,” in *Oxford Handbook of Economic Forecasting*, ed. by M. P. Clements, and D. F. Hendry, pp. 355–390. Oxford University Press.

Bao, Y., T.-H. Lee, and B. Saltoglu (2007): “Comparing Density Forecast Models,” *Journal of Forecasting*, 26, 203–225.

Boero, G., J. Smith, and K. F. Wallis (2008): “Uncertainty and Disagreement in Economic Prediction: The Bank of England Survey of External Forecasters,” *Economic Journal*, 118, 1107–1127.

——— (2010): “Scoring Rules and Survey Density Forecasts,” *International Journal of Forecasting*, 27, 379–393.

Bomberger, W. A. (1996): “Disagreement as a Measure of Uncertainty,” *Journal of Money, Credit and Banking*, 28, 381–392.

Carlson, J. A., and M. Parkin (1975): “Inflation Expectations,” *Economica*, 42, 123–138.

Cenesizoglu, T., and A. Timmermann (2008): “Is the Distribution of Stock Returns Predictable?,” Working Paper, University of California at San Diego.

Clements, M., and D. I. Harvey (2011): “Combining Probability Forecasts,” *International Journal of Forecasting*, 27, 208–223.

Coroneo, L., and D. Veredas (2010): “A Simple Two-Component Model for the Distribution of Intraday Returns,” *European Journal of Finance*, forthcoming.

DeMiguel, V., L. Garlappi, and R. Uppal (2009): “Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?,” *Review of Financial Studies*, 22(5), 1915–1953.

Dette, H., N. Neumeyer, and K. F. Pilz (2006): “A Simple Nonparametric Estimator of a Strictly Monotone Regression Function,” *Bernoulli*, 12, 469–490.

Dette, H., and S. Volgushev (2008): “Non-Crossing Non-Parametric Estimates of Quantile Curves,” *Journal of the Royal Statistical Society, Series B*, 70, 609–627.

Diebold, F. X., and R. S. Mariano (1995): “Comparing Predictive Accuracy,” *Journal of Business & Economic Statistics*, 13, 253–263.

Engelberg, J., C. F. Manski, and J. Williams (2009): "Comparing the Point Predictions and Subjective Probability Distributions of Professional Forecasters," *Journal of Business & Economic Statistics*, 27, 30–41.

Geweke, J., and G. Amisano (2011): "Optimal Prediction Pools," *Journal of Econometrics*, 164, 130–141.

Giordani, P., and P. Söderlind (2003): "Inflation Forecast Uncertainty," *European Economic Review*, 47, 1037–1059.

Gneiting, T., and A. E. Raftery (2007): "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association*, 102, 359–378.

Good, I. (1952): "Rational Decisions," *Journal of the Royal Statistical Society, Series B*, 14, 107–114.

Hall, S. G., and J. Mitchell (2007): "Combining Density Forecasts," *International Journal of Forecasting*, 23, 1–13.

Hansen, P. R. (2005): "A Test for Superior Predictive Ability," *Journal of Business & Economic Statistics*, 23, 365–380.

Hansen, P. R., and A. Lunde (2005): "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?," *Journal of Applied Econometrics*, 20, 873–889.

Härdle, W., M. Müller, S. Sperlich, and A. Werwatz (2004): *Nonparametric and Semiparametric Models*. Springer.

Jore, A. S., J. Mitchell, and S. P. Vahey (2010): "Combining Forecast Densities from VARs with Uncertain Instabilities," *Journal of Applied Econometrics*, 25, 621–634.

Jose, V. R. R., and R. L. Winkler (2008): "Simple Robust Averages of Forecasts: Some Empirical Results," *International Journal of Forecasting*, 24(1), 163 – 169.

Kascha, C., and F. Ravazzolo (2010): "Combining Inflation Density Forecasts," *Journal of Forecasting*, 29, 231–250.

Koenker, R. (2005): *Quantile Regression*. Cambridge University Press.

Komunjer, I. (2005): "Quasi-Maximum Likelihood Estimation for Conditional Quantiles," *Journal of Econometrics*, 128, 137–164.

Kullback, S., and R. A. Leibler (1951): "On Information and Sufficiency," *The Annals of Mathematical Statistics*, 22, 79–86.

Lahiri, K., and X. Sheng (2010): "Measuring Forecast Uncertainty by Disagreement: The Missing Link," *Journal of Applied Econometrics*, 25, 514–538.

Mankiw, N. G., R. Reis, and J. Wolfers (2003): "Disagreement about Inflation Expectations," *NBER Macroeconomics Annual*, 18, 209–248.

Manski, C. F. (2004): "Measuring Expectations," *Econometrica*, 72, 1329–1376.

Nolte, I., S. Nolte, and W. Pohlmeier (2010): "The Good, the Bad and the Ugly: Analyzing Forecasting Behavior within a Quantal Response Framework with Misclassification," Working Paper, Warwick Business School.

Nolte, I., and W. Pohlmeier (2007): "Using Forecasts of Forecasters to Forecast," *International Journal of Forecasting*, 23(1), 15 – 28.

Pesaran, M. H., and M. Weale (2006): "Survey Expectations," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. W. Granger, and A. Timmermann. Elsevier.

Rubinstein, R. Y., and D. P. Kroese (2008): *Simulation and the Monte Carlo Method*. Wiley-Interscience, 2 edn.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.

Smith, J., and K. F. Wallis (2009): "A Simple Explanation of the Forecast Combination Puzzle," *Oxford Bulletin of Economics and Statistics*, 71, 331–355.

Stock, J., and M. W. Watson (2002): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

Timmermann, A. (2006): "Forecast Combinations," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. W. Granger, and A. Timmermann. Elsevier.

Wallis, K. F. (2005): "Combining Density and Interval Forecasts: A Modest Proposal," *Oxford Bulletin of Economics and Statistics*, 67, 983–994.

Winkler, R. L. (1969): "Scoring Rules and the Evaluation of Probability Assessors," *Journal of the American Statistical Association*, 64, 1073–1078.

——— (1996): "Scoring Rules and the Evaluation of Probabilities," *TEST*, 5(1), 1–26.

Zarnowitz, V. A., and L. A. Lambros (1987): "Consensus and Uncertainty in Economic Prediction," *Journal of Political Economy*, 95, 591–621.

Data Appendix

Series name	Code	Description	Tf
Real Gross Domestic Product	GDPC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real Personal Consumption Expenditures	PCECC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real Personal Consumption Expenditures: Durable Goods	PCDGCC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real Personal Consumption Expenditures: Nondurable Goods	PCNDGC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real Personal Consumption Expenditures: Services	PCESVC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real Gross Private Domestic Investment	GPDIC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real Private Fixed Investment	FPIC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real Private Nonresidential Fixed Investment	PNFIC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real Nonresidential Investment: Equipment & Software	NRIPDC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real Private Residential Fixed Investment	PRFIC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real Exports of Goods & Services	EXPGSC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real Imports of Goods & Services	IMPGSC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real Government Consumption Expenditures & Gross Investment	GCEC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real Federal Consumption Expenditures & Gross Investment	FGCEC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Real State & Local Consumption Expenditures & Gross Investment	SLCEC96	Billions of Chained 2005 Dollars, Quarterly, Seasonally Adjusted Annual Rate	D
Industrial Production Index	INDPRO	Index 2002=100, Monthly, Seasonally Adjusted	D
Industrial Production: Final Products (Market Group)	IPFINAL	Index 2002=100, Monthly, Seasonally Adjusted	D
Industrial Production: Consumer Goods	IPCONGD	Index 2002=100, Monthly, Seasonally Adjusted	D
Industrial Production: Durable Consumer Goods	IPDCONGD	Index 2002=100, Monthly, Seasonally Adjusted	D
Industrial Production: Nondurable Consumer Goods	IPNCONGD	Index 2002=100, Monthly, Seasonally Adjusted	D
Industrial Production: Business Equipment	IPBUSEQ	Index 2002=100, Monthly, Seasonally Adjusted	D
Industrial Production: Materials	IPMAT	Index 2002=100, Monthly, Seasonally Adjusted	D
Industrial Production: Durable Materials	IPDMAT	Index 2002=100, Monthly, Seasonally Adjusted	D
Industrial Production: nondurable Materials	IPNMAT	Index 2002=100, Monthly, Seasonally Adjusted	D
All Employees: Total Private Industries	USPRIV	Thousands, Monthly, Seasonally Adjusted	D
All Employees: Goods-Producing Industries	USGOOD	Thousands, Monthly, Seasonally Adjusted	D
All Employees: Natural Resources & Mining	USMINE	Thousands, Monthly, Seasonally Adjusted	D
All Employees: Construction	USCONS	Thousands, Monthly, Seasonally Adjusted	D
All Employees: Durable Goods Manufacturing	DMANEMP	Thousands, Monthly, Seasonally Adjusted	D
All Employees: Nondurable Goods Manufacturing	NDMANEMP	Thousands, Monthly, Seasonally Adjusted	D
All Employees: Service-Providing Industries	SRVPRD	Thousands, Monthly, Seasonally Adjusted	D
All Employees: Trade, Transportation & Utilities	USTPU	Thousands, Monthly, Seasonally Adjusted	D
All Employees: Wholesale Trade	USWTRADE	Thousands, Monthly, Seasonally Adjusted	D
All Employees: Retail Trade	USTRADE	Thousands, Monthly, Seasonally Adjusted	D
All Employees: Financial Activities	USFIRE	Thousands, Monthly, Seasonally Adjusted	D
All Employees: Government	USGOVT	Thousands, Monthly, Seasonally Adjusted	D
Civilian Labor Force	CLF16OV	Thousands, Monthly, Seasonally Adjusted	D
Nonfarm Business Sector: Hours of All Persons	HOANBS	Index 1992=100, Quarterly, Seasonally Adjusted	D
Average Weekly Hours: Manufacturing	AWHMAN	Hours, Monthly, Seasonally Adjusted	A
Average Weekly Hours: Overtime: Manufacturing	AWOTMAN	Hours, Monthly, Seasonally Adjusted	C

Table 4: Series used for construction of the principal component pc_t appearing in models four and five above. All series have been downloaded from the FRED database administered by the Federal Reserve of St. Louis; the corresponding (FRED-internal) series codes are listed in the second column. Data transformations A - E ("Tf", fourth column) are defined in Table 5 below.

Series name	Code	Description	Tf
Civilian Unemployment Rate	UNRATE	Percent, Monthly, Seasonally Adjusted	C
Average (Mean) Duration of Unemployment	UEMPMEAN	Weeks, Monthly, Seasonally Adjusted	C
Civilians Unemployed - Less Than 5 Weeks	UEMPLT5	Thousands, Monthly, Seasonally Adjusted	D
Civilian Unemployed for 5-14 Weeks	UEMP5TO14	Thousands, Monthly, Seasonally Adjusted	D
Civilians Unemployed - 15 Weeks & Over	UEMP15OV	Thousands, Monthly, Seasonally Adjusted	D
Civilians Unemployed for 15-26 Weeks	UEMP15T26	Thousands, Monthly, Seasonally Adjusted	D
Civilians Unemployed for 27 Weeks and Over	UEMP27OV	Thousands, Monthly, Seasonally Adjusted	D
Civilian Participation Rate	CIVPART	Percent, Monthly, Seasonally Adjusted	C
Housing Starts: Total: New Privately Owned Housing Units Started	HOUST	Thousands of Units, Monthly, Seasonally Adjusted Annual Rate	B
New Private Housing Units Authorized by Building Permit	PERMIT	Thousands of Units, Monthly, Seasonally Adjusted Annual Rate	B
Housing Starts in Northeast Census Region	HOUSTNE	Thousands of Units, Monthly, Seasonally Adjusted Annual Rate	B
Housing Starts in Midwest Census Region	HOUSTMW	Thousands of Units, Monthly, Seasonally Adjusted Annual Rate	B
Housing Starts in South Census Region	HOUSTS	Thousands of Units, Monthly, Seasonally Adjusted Annual Rate	B
Housing Starts in West Census Region	HOUSTW	Thousands of Units, Monthly, Seasonally Adjusted Annual Rate	B
ISM Manufacturing: PMI Composite Index	NAPM	Index, Monthly, Seasonally Adjusted	A
ISM Manufacturing: New Orders Index	NAPMNOI	Index, Monthly, Seasonally Adjusted	A
ISM Manufacturing: Supplier Deliveries Index	NAPMSDI	Index, Monthly, Seasonally Adjusted	A
ISM Manufacturing: Inventories Index	NAPMII	Index, Monthly, Seasonally Adjusted	A
Gross Domestic Product: Chain-type Price Index	GDPCTPI	Index 2005=100, Quarterly, Seasonally Adjusted	E
Personal Consumption Expenditures: Chain-type Price Index	PCECTPI	Index 2005=100, Quarterly, Seasonally Adjusted	E
Consumer Price Index For All Urban Consumers: All Items	CPIAUCSL	Index 1982-84=100, Monthly, Seasonally Adjusted	E
Personal Consumption Expenditures: Chain-Type Price Index Less Food and Energy	PCEPILFE	Index 2005=100, Monthly, Seasonally Adjusted	E
Consumer Price Index for All Urban Consumers: All Items Less Food & Energy	CPILFESL	Index 1982-84=100, Monthly, Seasonally Adjusted	E
Consumer Price Index for All Urban Consumers: Food	CPIUFDSL	Index 1982-84=100, Monthly, Seasonally Adjusted	E
Consumer Price Index for All Urban Consumers: Apparel	CPIAPPSL	Index 1982-84=100, Monthly, Seasonally Adjusted	E
Consumer Price Index for All Urban Consumers: Energy	CPIENGSL	Index 1982-84=100, Monthly, Seasonally Adjusted	E
Consumer Price Index for All Urban Consumers: Transportation	CPITRNSL	Index 1982-84=100, Monthly, Seasonally Adjusted	E
Consumer Price Index for All Urban Consumers: Medical Care	CPIMEDSL	Index 1982-84=100, Monthly, Seasonally Adjusted	E
Gross Private Domestic Investment: Chain-type Price Index	GPDICTPI	Index 2005=100, Quarterly, Seasonally Adjusted	E
Spot Oil Price: West Texas Intermediate	OILPRICE	Dollars per Barrel, Monthly	D
Average Hourly Earnings: Construction	AHECONS	Dollars per Hour, Monthly	D
Average Hourly Earnings: Manufacturing	AHEMAN	Dollars per Hour, Monthly	D
Business Sector: Output Per Hour of All Persons	OPHPBS	Index 1992=100, Quarterly, Seasonally Adjusted	D
Nonfarm Business Sector: Real Compensation Per Hour	COMPRNFB	Index 1992=100, Quarterly, Seasonally Adjusted	D
Nonfarm Business Sector: Unit Labor Cost	ULCNFB	Index 1992=100, Quarterly, Seasonally Adjusted	D
Effective Federal Funds Rate	FF	Percent, Weekly Ending Wednesday	C
3-Month Treasury Bill: Secondary Market Rate	WTB3MS	Percent,Weekly Ending Friday	C
5-Year Treasury Constant Maturity Rate	WGS5YR	Percent,Weekly Ending Friday	C
10-Year Treasury Constant Maturity Rate	WGS10YR	Percent,Weekly Ending Friday	C

Table 4 (cont'd): Series used for construction of the principal component pc_t appearing in models four and five above. All series have been downloaded from the FRED database administered by the Federal Reserve of St. Louis; the corresponding (FRED-internal) series codes are listed in the second column. Data transformations A - E ("Tf", fourth column) are defined in Table 5 below.

Series name	Code	Description	Tf
Moody's Seasoned Aaa Corporate Bond Yield	WAAA	Percent,Weekly Ending Friday	C
Moody's Seasoned Baa Corporate Bond Yield	WBAA	Percent,Weekly Ending Friday	C
WGS10YR - WTB3MS	-	Percent,Weekly Ending Friday	A
WAAA - WGS10YR	-	Percent,Weekly Ending Friday	A
WBAA - WGS10YR	-	Percent,Weekly Ending Friday	A
Dow Jones Industrial*	S19655	Index, Daily	D
Consumer Confidence*	440005021	Index, Monthly, Seasonally Adjusted	C
Real Estate Loans at All Commercial Banks	REALLN	Billions of Dollars, Monthly, Seasonally Adjusted	D
Consumer (Individual) Loans at All Commercial Banks	CONSUMER	Billions of Dollars, Monthly, Seasonally Adjusted	D
Commercial and Industrial Loans at All Commercial Banks	BUSLOANS	Billions of Dollars, Monthly, Seasonally Adjusted	D
M1 Money Stock	M1SL	Billions of Dollars, Monthly, Seasonally Adjusted	E
M2 Money Stock	M2SL	Billions of Dollars, Monthly, Seasonally Adjusted	E
Bank Prime Loan Rate	MPRIME	Percent, Monthly	C

Table 4 (cont'd): Series used for construction of the principal component pc_t appearing in models four and five above. Series not marked with * have been downloaded from the FRED database administered by the Federal Reserve of St. Louis; the corresponding (FRED-internal) series codes are listed in the second column. Series marked with * have been downloaded from data stream; here the second column displays the data stream series code. Data transformations A - E ("Tf", fourth column) are defined in Table 5 below.

Code	Transformation (Tf)
A	Y_t
B	$\ln(Y_t)$
C	$Y_t - Y_{t-1}$
D	$\ln(Y_t) - \ln(Y_{t-1})$
E	$\ln\left(\frac{Y_t}{Y_{t-1}}\right) - \ln\left(\frac{Y_{t-1}}{Y_{t-2}}\right)$

Table 5: Data Transformations A - E used in Table 4 above. Y_t denotes the original value of the series.